

Executive Summary: After noticing that the Minecraft speedrunner Dream had unusually good luck in a subset of livestream runs, the question arose about whether he modified certain parts of the game. Determining the true odds of "lucky streaks" after they occur requires detailed statistical analysis. The Minecraft Speedrunning Team produced an official report that claimed the highest possible odds of Dream's results were 1 in 7.5 trillion even when correcting for biases. Commissioned by Dream, I perform a review of this original report and a second expert statistical analysis that I argue is more accurate. The odds are about 1 in 10 million that a small subset of any livestreamed speedruns from any player in the past year would give as low a probability if investigated in any two ways – if only the six "lucky" streams are investigated. The higher odds in my analysis result from a higher fidelity simulation of when speedrunners stop bartering and an improved correction for some of the biases. If all eleven streams discussed are included, then the low probability events are consistent with random chance. Deciding between these odds depends on external considerations, but it is much too extreme to state that there is a 1 in 7.5 trillion chance that Dream did not cheat.

Abstract

I study the Minecraft Speedrunning Team (MST) Report investigating the speedrunner known as Dream. The MST Report concludes that Dream modified his runs which is denied by Dream. **Dream commissioned this independent analysis to get a second expert opinion**, though he did not directly influence it. I identify two major issues with the MST Report: it does not account for stopping bartering after a successful trade and it incorrectly applies some bias corrections. An independent analysis using my best estimates, Bayesian statistics, and bias corrections gives a higher probability of about 1 in 100 million that any Minecraft speedrunner would have experienced two sets of improbable events during the past year like Dream did if the game was modified before the six final streams. The two main reasons for the higher odds are 1) a higher fidelity accounting for "barter stopping" after getting 10 ender pearls (factor of about 100) and 2) a more accurate estimate of the number of potentially investigated random aspects and the number of meaningful livestream speedrun comparisons. That it was Dream, specifically, who experienced this extremely rare event is already accounted for by the fact that he was investigated because his streams seemed improbable; comparison to the records of other speedrunners should not be considered independent evidence. The MST Report hypothesizes that Dream's return to speedrunning prompted a modification and thus considers the six post-return streams alone. Five previous streams were consistent with default probabilities. If these are included in the analysis and the bias corrections applied, there is no significant evidence that the game was modified. Determining which probability is most appropriate requires assessing the odds – independent of the outcomes of the streams – comparing whether Dream would have made a modification at the beginning of all eleven streams versus the beginning of the final six streams. An attempt to correct for the bias that any subset could have been considered changes the probability of Dream's results to 1 in 10 million or better. The probabilities are not so extreme as to completely rule out any chance that Dream used the unmodified probabilities. However, the probability of the hypothesis that the game was modified in two ways before his final six runs is quite low even when correcting for bias. Although this could be due to extreme "luck", the low probability suggests an alternative explanation may be more plausible. One obvious possibility is that Dream (intentionally or unintentionally) cheated. Assessing this probability exactly depends on the range of alternative explanations that are entertained which is beyond the scope of this document, but it can depend highly on the probability (ignoring the probabilities) that Dream decided to modify his runs in between the fifth and sixth (of 11) livestreams. This is a natural breaking point, so this hypothesis is plausible. In any case, the conclusion of the MST Report that there is, at best, a 1 in 7.5 trillion chance that Dream did not cheat is too extreme for multiple reasons discussed herein.

Critique of Dream Investigation Results

Photoexcitation

December 21, 2020

Contents

1	What is this report?	2
2	Who wrote this document?	3
3	What are the goals of this document?	3
4	Statistics Prelude	4
4.1	Statistical Modeling	4
4.2	Hypothesis Testing vs Bayesian Modeling	5
5	Context for Statistical Analysis	5
6	Inappropriateness of the Binomial Distribution	6
6.1	Binomial Distribution for Blaze Rod Drops	9
6.2	Probability Evaluations for Ender Pearls	9
6.3	Stopping Criterion	10
6.4	Blaze Rod Probability	10
6.5	Joint Probability	12
7	Inappropriate Correction for Sampling Bias	12
7.1	Inaccurate Correction for Lucky Streaks	12
7.2	Actual Probability of Lucky Streaks	13
7.3	Including all 11 streams	13
8	Other Corrections	14
8.1	How many random events are important?	14
8.2	Combined Correction	15
8.3	Comparing to Other Speedrunners	15
9	Conclusions	15
A	Code Snippets	17

1 What is this report?

This report is a discussion of the Minecraft Speedrunner "Dream"¹ who, during livestreams with speedrunning attempts of the Minecraft 1.16 Random Seed Glitchless, experienced very low probability events over a seemingly specific length of time. Extremely rare events pique our interest and can require an explanation, e.g., for the purpose of deciding whether Dream's speedruns are eligible for official leaderboards.

¹<https://www.twitch.tv/dreamwastaken>

After taking note of the very unlikely events, the Minecraft Speedrunning Team (MST) wrote an official Report (hereafter "MST Report" available at <https://mcspeedrun.com/dream.pdf>) into some of Dream's streamed speedruns. The final sentence of the MST Report is "the only sensible conclusion that can be drawn after this analysis is that Dream's game was modified in order to manipulate the pearl barter and rod drop rates." A related less-formal YouTube video explains some of these details. Dream has claimed that he did not (intentionally) modify his game, although conclusive evidence of this may be impossible to obtain.

The MST Report attempted to account for possible bias in multiple ways, emphasizing their desire to be as favorable as possible to Dream. This document attempts to explain some major concerns about the statistical methods used in the MST Report. Addressing these concerns would make the probability that Dream did not cheat substantially more favorable, although I do not repeat the MST Report analysis with an improved methodology. This report also provides an independent statistical analysis. **This report was commissioned by Dream, but he did not have direct or undue influence over the outcome.** For example, Dream provided feedback on this report, but was not an author of any portion of it.

2 Who wrote this document?

This article was written by an expert from the online science consulting company Photoexcitation (see <https://www.photoexcitation.com/>). As with all Photoexcitation activities, the exact identity of the author will not be revealed. Similarly to the MST Report, arguably the authorship does not matter because the analysis is intended to be objective and verifiable by anyone with sufficient background. However, it is helpful to discuss some key details about the authorship.

There was only one author and for simplicity in explanation, I will use first-person pronouns.

First, it is imperative to **disclose that this report was sought out and commissioned by Dream.** Despite this financial backing, I did not focus any effort on exonerating Dream and express my opinion that Dream himself was primarily interested in a second expert opinion. One top goal was to provide a rebuttal – where objectively justifiable – to the MST Report. This document can be seen as similar to a "referee report" provided by scientists in the peer-reviewed journal literature.

I am an active practicing astrophysicist who is regularly called upon by journals, federal grant review panels, colleagues, clients, and others to provide extensive feedback. I have extensive expertise in statistics, having multiple direct connections to the field of astrostatistics² I am fully expert in statistics at the level required to provide objective, meaningful, and accurate feedback. I was vaguely familiar with Minecraft and am now more familiar after researching and writing this report. I used as primary sources the MST Report, discussions with Dream, some online comments, someone who emailed Dream who wishes to remain anonymous, and my own experience. The MST Report explains its reasoning very well even to non-Minecraft experts and provided most of the key information. For example, I use the same data as is listed in their Appendix A.

Dream commissioned this report and provided direct feedback, but was not a coauthor.

3 What are the goals of this document?

The goal of this document is to discuss the probability calculations performed in the MST Report and to provide a second opinion. There is no explicit goal to exonerate Dream or to reach a more favorable conclusion.

This document does **not** have the goal of:

1. Arguing that Dream's speedrun should be reinstated. Dream has expressed to me that he is not concerned about his leaderboard position and is more concerned about the perception of his character.
2. Providing evidence or speculation that the MST Report was biased, although its accuracy is assessed.
3. Investigating the MST Report's discussion of Code Analysis (their Section 9). Though the author is expert in these aspects of code analysis as well, looking into this was not a goal of this document. A

²Yes, "astrostatistics" is a real field, see <https://asaip.psu.edu/>.

brief perusal suggests that this section is accurate. I will assume that numbers that are supposed to be random *are* truly random.

The author's opinion is that the MST Report was well-written and was mostly correct in how it assessed Dream's odds. It provided an explanation that works well for both the layman and the expert. However, there are several issues and inaccuracies that are addressed here.

4 Statistics Prelude

Some initial discussion of statistics will be helpful as a prelude. I will not be reviewing the basic statistical analysis information from Section 7 of the MST Report, so if you are unfamiliar the basics of probability and statistics, you may wish to start there.

As with all objective and scientific statistical analyses, I assume as an axiom that there is no such thing as luck. Luck is just a concept that we associate with low probability events. However, it is sometimes useful to communicate the ideas of probability in terms of "luckiness" or "unluckiness" and I will do so in this document.

Another important concept to remember (in this report and in life) is that *one in a billion events happen every day*. People win the lottery... some win the lottery multiple times! Just because an event is rare, even surprisingly rare, does not mean it should be rejected.

The goal of computing probabilities is to allow us to draw conclusions and make decisions. Maybe your friend will decide to believe Dream if the probability is one in a billion, but you need the odds to be "only" one in a million before you'll side with Dream. As a result, some of the responsibility for interpretation falls to the reader.

4.1 Statistical Modeling

Probability calculations are hard. There may not be one "right" way to do something. It is easy to violate some hidden or unknown assumption. There is room for healthy debate about different methods and results.

A gold standard method in statistical analysis is known as "forward modeling" which is using a simulation of an event to study probabilities. The appropriateness and accuracy of forward modeling as a method is very difficult to question. Instead, the question should be about the *fidelity* of the forward model: how accurately does it describe the situation which actually lead to the observed data? In practice, this is usually handled by comparing two different models by assessing which one is higher fidelity. Ideally, competing forward models are both run to see 1) if there is a difference and 2) if the difference makes sense. When thinking about forward models, it is also important to remember a common statistics adage: "all models are wrong, but some are useful." There will always be a way to improve a model's approximation to reality (e.g., all models are wrong), but at some point you reach a fidelity that is considered acceptable and appropriate (e.g., a useful model).

Most of the rules and laws and methods of statistics are basically shortcuts to the full forward modeling process, e.g., using mathematical equations to run a precise or approximate "simulation." Some approximations are better than others, of course. Many approximations have hidden or unwritten assumptions that can be violated unintentionally, leading to an inaccurate result.

In the process of developing forward models of higher and higher fidelity, one disadvantage becomes computational tractability. Some forward models take so long that they can't be completed without unreasonable computational time. Some of the models in this document took about an hour to complete on a modern machine and others were not even considered due to their complexity.

For assessing probability, a common forward modeling technique is known as Monte Carlo method³. In this method a large number of simulations are generated using random numbers. Some interesting property of these numbers (sometimes called a "statistic") is then calculated. By comparing the distribution of this "statistic" with what is seen in the actual data, a probability (called a "p-value") can be assessed. A "p-value" can be interpreted as the probability that an event would happen by random chance. One important aspect to remember about Monte Carlo simulations is that they are based on random samples and so do

³see, e.g., https://en.wikipedia.org/wiki/Monte_Carlo_method#Applied_statistics

have some variation from simulation to simulation. The size scale of this variation is typically the square root of the number of successes and thus it is typical to use 10^5 - 10^7 simulations, like I do in my analysis. For example, if a certain value of a statistic occurs 9 times in 10^6 simulations, then the uncertainty on this result can be approximated by $\sqrt{9} = 3$, e.g., the p-value would be $9 \pm 3 \times 10^{-6}$. Higher precision can be obtained by running much larger simulations. In this document, my goal is to get factor-of-a-two precision. That is, odds of "1 in 4 billion" should not be interpreted as substantively different than "1 in 2 billion" or "1 in 8 billion", but is different from "1 in 20 billion". Combining this notion with the standard scientific practice of communicating higher precision by using more digits ("significant figures"), my estimates will typically only be listed to 1-2 digits of precision.

4.2 Hypothesis Testing vs Bayesian Modeling

Although not explicitly written this way, the MST Report typically focuses on a hypothesis testing paradigm for its calculations. That is, they propose a null hypothesis, "Dream had unmodified probabilities in all of his runs" and then attempt to reject this hypothesis by calculating "p-values" (probabilities for null hypothesis rejection under certain assumptions).

Another probabilistic paradigm is Bayesian statistics. Instead of comparing the data to a random set, it compares the relative probabilities of different choices for "model parameters." For example, I use a Bayesian model where the parameter is "by what factor was the ender pearl probability enhanced?" and use this to consider the probability of the unenhanced case. This parameter is chosen because it naturally distinguishes between the unmodified and modified cases. Choosing this as a parameter does not imply that the probabilities were enhanced.

In this document, I don't have time to discuss the long-term debate between these different statistical paradigms and when they should be applied. The short version is that another way of investigating whether the probabilities were modified is to try to determine what probabilities were used. The probability of a particular probability enhancement (including no enhancement) can then be calculated.

5 Context for Statistical Analysis

After careful reading of the MST Report and correspondence with Dream, it is important to clearly identify what I am investigating.

Like hundreds of other speedrunners, Dream plays and livestreams regularly with various goals and multiple versions of Minecraft. The extensive amounts of data gathering that would be required to monitor all possible cases of inappropriate modifications is intractable. As a result, investigations are only triggered if it seems like someone experiences a beneficial very low probability event. Within these investigations, basic data can then be gathered but only for specific aspects of specific streams, with a particular focus on what seems most unusual.

Extremely low probability events regularly happen. If you consider every Minecraft player, then a "perfect" ender pearl and blaze drop record (2/2 ender pearl barbers and 7/7 blaze rod drops) occurs *multiple times per hour*, since this has a 1 in 60000 odds and Minecraft is played many millions of times a day. Considering all Minecraft worlds ever played and the multitude of ways in which luck plays a role, even one in a trillion events happen daily.

Of course, the vast majority of these events happen off camera and under no scrutiny. Experiencing a rare event – like the perfect run above – and then reporting it on twitter would not be surprising. This is reminiscent of a story of Richard Feynman – brilliant physicist of the mid 20th century – who was pointing out a probability fallacy. He is quoted as saying

You know, the most amazing thing happened to me tonight. I was coming here, on the way to the lecture, and I came in through the parking lot. And you won't believe what happened. I saw a car with the license plate ARW 357. Can you imagine? Of all the millions of license plates in the state, what was the chance that I would see that particular one tonight? Amazing!

In his usual pedagogical way, Feynman used sarcasm to illustrate a point to teach scientists about the crucial importance of skepticism. Of course, this situation was not amazing or unusual because you can replace

"ARW 357" with *any* licence plate and say the same thing. The key point here is that it is not unlikely to identify an improbable event *after* it happens. But it *is* unlikely to predict an improbable event in advance. For example, if I say "the next license plate you see will be WPB 162", I would need to be pretty lucky for that to turn out to be true because I predicted a specific unlikely sequence *in advance*. (Although if millions of people read this document, one of them probably *would* see WPB 162 first!).

So, a major challenge of investigating Dream's record is that any series of streams that is scrutinized precisely because it seems unusual introduces a strong bias. This is known as "cherry picking" and is a legitimate concern in any analysis of events triggered because they are unlikely. As the MST Report states, it is possible to correct for this bias and estimate the probability despite only choosing to investigate unusual events. To be clear, this is not a question of whether the MST were objective or had a hidden agenda (for or against Dream), although those can also influence their choice in which investigations to pursue which can potentially factor in to the how the resulting probability should be interpreted. For the purposes of this document, I make no assumptions or assertions about MST's motives other than their self-admitted choice of investigating a specific set of runs precisely because they were unusually low probability.

The number of interacting variables and components is too complex to come down to a single answer for "this is the probability that Dream modified his streams". Thus, a goal of the MST Report was to identify and attempt to study and mitigate the strongest potential biases. They focus on the following:

1. The non-binomial nature of the probability of events that have a result-based stopping criterion.
2. They would have investigated reports of this low of a probability or lower, so cumulative binomial probabilities should be considered (a common choice for hypothesis testing).
3. They could have investigated any subset of consecutive streams and chose a specific set of six from eleven because those six had low probabilities.
4. They could have investigated any of about 1000 speedrunners, but only investigated this case because it was unusual.
5. They could have investigated a variety of possible aspects of these runs, but chose to investigate ender pearls because that seemed to be where the probabilities were modified. Blaze drops were added later to the investigation because of their connection to ender pearls and their seeming low probability.

The strength of the MST Report is its claim that, despite giving Dream the benefit of the doubt in many of these areas – which increased the raw probability by a factor of about 10 million (see Equations 11 and 16) – the probability of an unmodified run was still extremely low (about 1 in 7.5 trillion).

I criticize here some of the methods used and some of the conclusions reached by the MST Report. My criticisms include

1. Ender pearl barterers should not be modeled with a binomial distribution because the last barter is not independent and identical to the other barterers.
2. Their method for correcting p-values based on the number of consecutive streams selected is not appropriate.
3. They did not always use appropriate statistics that are designed specifically for looking at unusual events.
4. Their method for identifying comparable runs for investigation was arguably too restrictive, leading to lower odds.

These and other issues will be discussed in detail.

6 Inappropriateness of the Binomial Distribution

In order to calculate an accurate probability, we need to use a model (whether mathematical or Monte Carlo) that captures as much of the actual process as possible. Let's consider now the case of gathering ender pearls

through piglin bartering. The MST Report proposes a model that each barter is fully independent and uses a binomial model to then calculate probabilities aggregated across runs.

However, in practice, Dream and other speedrunners will barter with piglins until they reach the desired number of ender pearls (typically 10-12) and will then immediately stop, leaving uncompleted other barterers. Outside the official report, there has been some discussion on how to appropriately account for this. On the one hand, care must be taken to avoid the Gambler's Fallacy that being unlucky in one area makes you more lucky in (an independent) area. Thus any "off-the-camera" unluckiness actually has no bearing on another chosen barter. For example, the fact that the previous barter was an ender pearl doesn't affect the probability on the next barter.

But the fact that a previous barter was an ender pearl *does* effect how many barterers are made. Thus comparing the number of pearls to the number of barterers can be affected by the outcome of other barterers. If the last barter in a sequence is always an ender pearl (because then the speedrunner leaves), then it simply cannot be claimed that all barterers are fully independent and identical. Without identical independent barterers, the binomial model is inappropriate.

Consider two simulations of what happens during speedrun bartering:

- "Barter Stopping" Simulation - stop bartering after receiving 10-12 ender pearls
- "Binomial" Simulation - every barter is identical and the number of barterers made is independent of the outcome of other barterers

The Binomial Simulation is so called because it is well modeled by the binomial distribution. However, the Barter Stopping Simulation is more accurate to what really happens in speedrun bartering. Neither model is perfect, but the Barter Stopping Simulation is higher fidelity and thus more useful than the Binomial Simulation.

One way of describing this difference is as a "stopping criterion" for bartering. The MST Report (Section 8.1 and Appendix B) discuss optional stopping, but that is focused entirely on Dream stopping after his final successful run, not on stopping within each bartering session. That separate stopping criterion is discussed below. For clarity, I refer to stopping after receiving 10-12 ender pearls within a bartering session as "Barter Stopping."

Comparing the two simulations shows that they do give different results when considering ender pearls. The code for the non-trivial simulations is given below and the results are shown graphically in Figure 2. The Barter Stopping Simulation does indeed show that there are fewer barterers required to get the desired number of ender pearls, as your intuition would tell you if you end immediately after an ender pearl barter. It also helps explain why charts showing Dream's bartering outcomes seem unbalanced with respect to ender pearls... they don't account for the fact that ender pearls are special because they are the explicit and pre-determined goal of bartering.

I have an approximate model for the number of pearls given (see code snippet below) that matches the observed distribution and was suggested by a contributor who wishes to remain anonymous. Variations in this model were not significant. In this model, a random value between 4-7 pearls are given with equal probability of each. To reach 10 ender pearls requires 2 barterers 81% of the time and 3 barterers the other 19%. When the goal is to reach 12 ender pearls, this takes 2 barterers 60% of the time and 3 barterers 40% of the time. Since my simulation always ends on a successful pearl barter, the probability can be computed using the rules of probability. For example, for the 12 pearl case, I have confirmed that my Monte Carlo simulation distribution gives the expected result of 0.6 times the binomial distribution of 1 successful barter plus 0.4 times the binomial distribution of 2 successful barterers, all multiplied by the probability of an ender pearl barter (the last barter).

Accounting for Barter Stopping – which the author considers to be objectively a higher fidelity and thus more accurate simulation – makes Dream's odds less extreme. However, even with Barter Stopping, Dream seems particularly lucky, since the typical number of barterers needed is about 20 and Dream's 22 trading sessions for the six streams in question are almost always better than this.

Note that these simulations only account for barterers where the goal is to get to a specific number of ender pearls. The Barter Stopping Simulation is not appropriate when

1. the goal of bartering is not to obtain 10 ender pearls

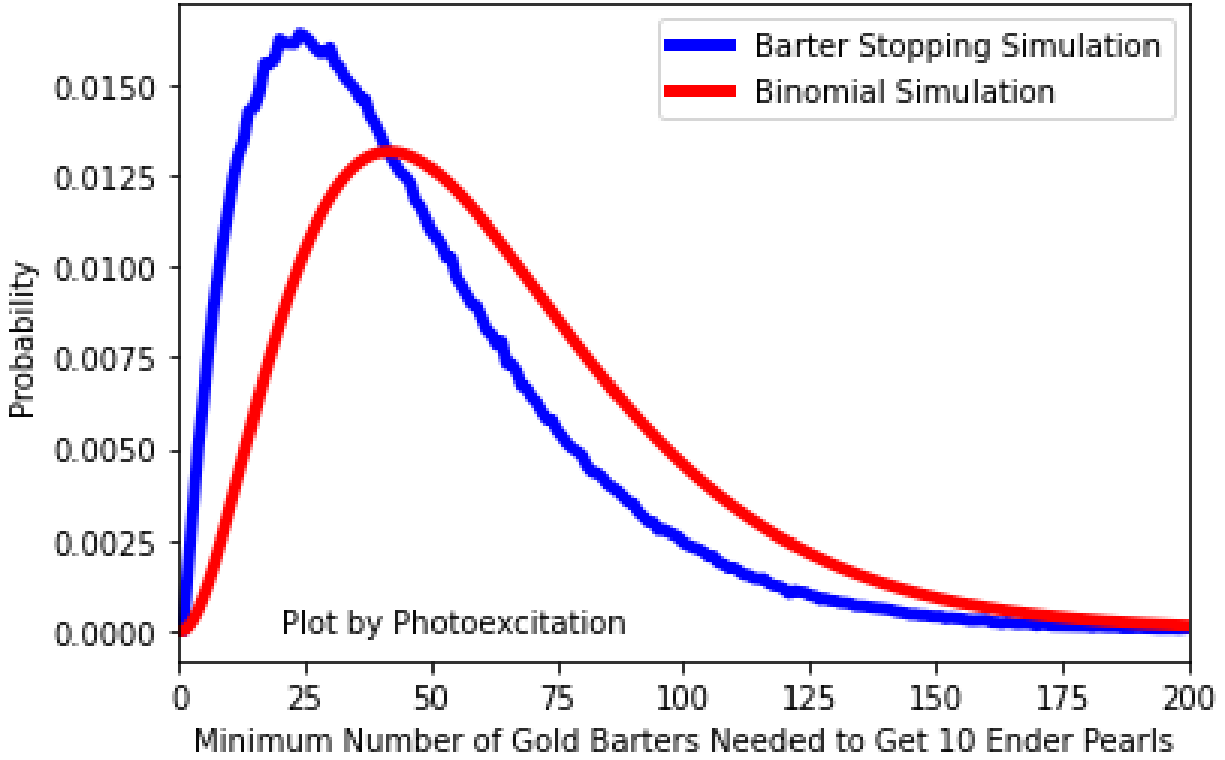


Figure 1: Comparison of two Simulations of piglin bartering for ender pearls. In the blue "Barter Stopping Simulation", gold barbers stop immediately after receiving 10 ender pearls. In the red "Binomial Simulation", every barter, including the last, is completely independent. In both cases, the x-axis represents the amount of gold and the y-axis represents successfully reaching 10 ender pearls (or 2 barbers in the Binomial case). The Barter Stopping Simulation is a more accurate reproduction of what speedrunners actually do when bartering for ender pearls. As can be seen, the typical number of gold barbers is lower in the Barter Stopping Simulation. Using the Binominal Simulation to assess the probability of ender pearl barbers makes speedrunners look more "lucky" than they actually are when barbers are conducted until 10 ender pearls are reached. When the goal of 12 ender pearls is used, the difference is weaker, but still significant.

2. 10 ender pearls are not obtained in a bartering session
3. bartering continues well after successful barterers
4. considering multiple separate attempts to obtain 10 ender pearls (e.g., combining multiple lines of data from MST Report Appendix A).

Without additional information on the motivations and context of barterers that fall into these four categories, the Binomial Model is a good approximation for the probability in these cases. For Dream, I will assume that his goal was always to obtain 10 ender pearls, so the Binomial Model is only used when 10 pearls were not obtained and when the bartering continued beyond 10 pearls.

I also considered a case where the goal was 12 ender pearls and which had associated probabilities for Dream's streams about 10 times lower. This makes sense because the Barter Stopping simulation emphasizes probabilities for fewer gold barterers. As can be seen from the distribution of ender pearls gathered, Dream rarely continued trading once 10 pearls were obtained. The 10-pearl probabilities are thus more appropriate for the simulation.

6.1 Binomial Distribution for Blaze Rod Drops

I can apply the same reasoning as above for blaze rod drops. Blaze sessions often continue until 7 blaze rods are obtained and then the speedrunner will continue on. Therefore, the last blaze drop is likely to be successful and thus not fully independent of the others, making the Binomial Model inappropriate.

When I simulate this process in the same way as with ender pearls, I find that the probabilities are not significantly different between a "Blaze Drop Stopping Simulation" and a "Binomial Simulation." There are two things that both make the blaze drop situation much better approximated by the binomial distribution. First, 6/7 blaze rod drops are independent (because they aren't the last drop) unlike the 1/2 cases for ender pearls. Second, the blaze rod drop probability of 0.5 is much higher than the $\frac{20}{423} \approx 0.0473$ for ender pearls, so the imbalance caused by "the last drop is the one I was looking for" is much less important. As a result, it is unsurprising that the "Blaze Drop Stopping" Simulation was not significantly different than the Binomial Simulation. In my calculations, I use the simpler binomial probabilities.

6.2 Probability Evaluations for Ender Pearls

I can now calculate the probability of receiving the number of successful trades for Dream's six streams in question. I use the Barter Stopping probabilities when 10 ender pearls were reached and Binomial probabilities when they were not.

Note that there is one case at the end of the second stream where 12 gold are bartered for 5 sets of ender pearls. Some data collection on this suggests that 4 sets of ender pearls were obtained. Either way, this is obviously a low probability event in any case and assigning this to Barter Stopping vs. Binomial probabilities can make an order of magnitude difference in the result. Arguably, Barter Stopping should not apply to this case, even though it is possible that Dream stopped trading once he was successful. This case was thus modeled with the Binomial probability.

Unfortunately, the use of the Barter Stopping probabilities makes the calculation of the overall probability more complicated. No longer can all the bartering sessions be combined into a single calculation. This then requires calculating probabilities not only for the each individual trading session, but also the distribution of golds used/available among the trading sessions. This quickly leads to a challenge in calculating probabilities that even simulations can't effectively get around, although this was attempted.

For these reasons and other reasons mentioned above, I choose to model the probability with Bayesian statistics. There are many arguments in the statistics literature that support using Bayesian statistics for calculating low probability events like this.

Within a Bayesian model, instead of calculating a probability that Dream did not use modifications, I instead compare the probability of different possible modifications. By comparing a range (1-5) of possible "ender pearl probability boosts", I can assess the probability that the probability boost/increase was equal to 1.0, e.g., the probabilities are the default Minecraft probabilities. The choice to use a parameter for ender pearl modifications reflects a desire to understand the probability that the observed data would occur and has no implication that the probabilities were modified.

Since Bayesian probability calculations are relative, constant factors (like the number of ways to partition the total number of trades into the specific observed data) cancel out. In particular, I follow the usual Bayesian technique and calculate the probability of getting exactly the observed data, i.e., using the Binomial probability mass function instead of the cumulative distribution function (which is used in non-Bayesian methods). This allows me to focus on the relative posterior probability of different boosts, with the probability of boost=1 compared to all other cases representing the probability that there were no modifications.

(For those savvy in Bayesian statistics, I use a flat/uniform/tophat prior on the probability boost from 1 to 5 and confirmed that these limits do not significantly affect the interpretation. In this case, this just means calculating the likelihoods on a grid from 1 to 5 and, since the prior is flat, these are equivalent to the relative posterior probabilities. This prior does not include any corrections for biases or any opinion that Dream modified his probabilities.)

Applying this technique to the observed data from the six streams results in a posterior distribution that is highly peaked around a probability boost of 3. At boost=1, the default case that Dream is arguing, the probability is only 3×10^{-10} . Doing a quick check on the case where only the Binomial Probability is used gives 5×10^{-12} . This reduction in probability by a factor of 100 is sensible given how much the Barter Stopping Model favors probabilities at low numbers of gold as seen in these streams. These probabilities are also similar to the probability estimated by the MST Report, with the most direct comparison to their naive estimate of 5.65×10^{-12} . As expected, using the Barter Stopping criterion increases the probability, though some of the difference may be attributable to the Bayesian modeling method as well.

However, this probability does not account for the fact that these streams were chosen for investigation specifically because they seemed low probability. That is, 3×10^{-10} is *not* the probability that Dream modified the ender pearl probabilities.

6.3 Stopping Criterion

There are many possible ways of considering and implementing stopping criteria. The main challenge is that once a speedrunner gets particularly lucky, they are more likely choose to stop playing. Dream has expressed that this was his stopping criterion. Indeed, Dream's final run was exceptionally fortuitous with only 3 gold barterers needed to get 2 ender pearl barterers. Since a speedrunner's final run tends to be low probability, a correction needs to be applied. The MST Report uses a detailed stopping algorithm to identify any combination of trades that gives an exceptionally low p-value and allows for stopping to occur in any of these cases. This is a reasonable approximation.

Implementing this particular stopping criterion is not practical with my setup. Instead, I propose a simpler case: drop the last datapoint. This removes the bulk of the issue since the speedrunner cannot know in advance that the next run will be lucky and thus the second-to-last run is effectively identical to all the other runs. Removing the final datapoint gives a Bayesian probability that there was no modification at 3×10^{-9} , about ten times better than when the last datapoint is kept. This is sensible as the last run was unusually successful. This stopping criterion removes another case that is unusually lucky from the data and may thus inappropriately increase the probability. For the sake of having a single concrete number, I choose to split the difference and use a probability of 10^{-10} as the chance that there were no ender pearl modifications in Dream's last six streams.

6.4 Blaze Rod Probability

Recall that the "Blaze Rod Drop Stopping" case was effectively the same as the Binomial case. Evaluating both using my Bayesian probabilistic method gives the same answer of 3×10^{-8} . The peak for the blaze rod probability (which was evaluated over a prior from 0.5 to 0.9, with limits that don't affect the answer) is around 0.7.

Removing the last blaze rod drop, which was favorable, from the list of 32 cases did not make a significant difference in the probability, so I use the above value.

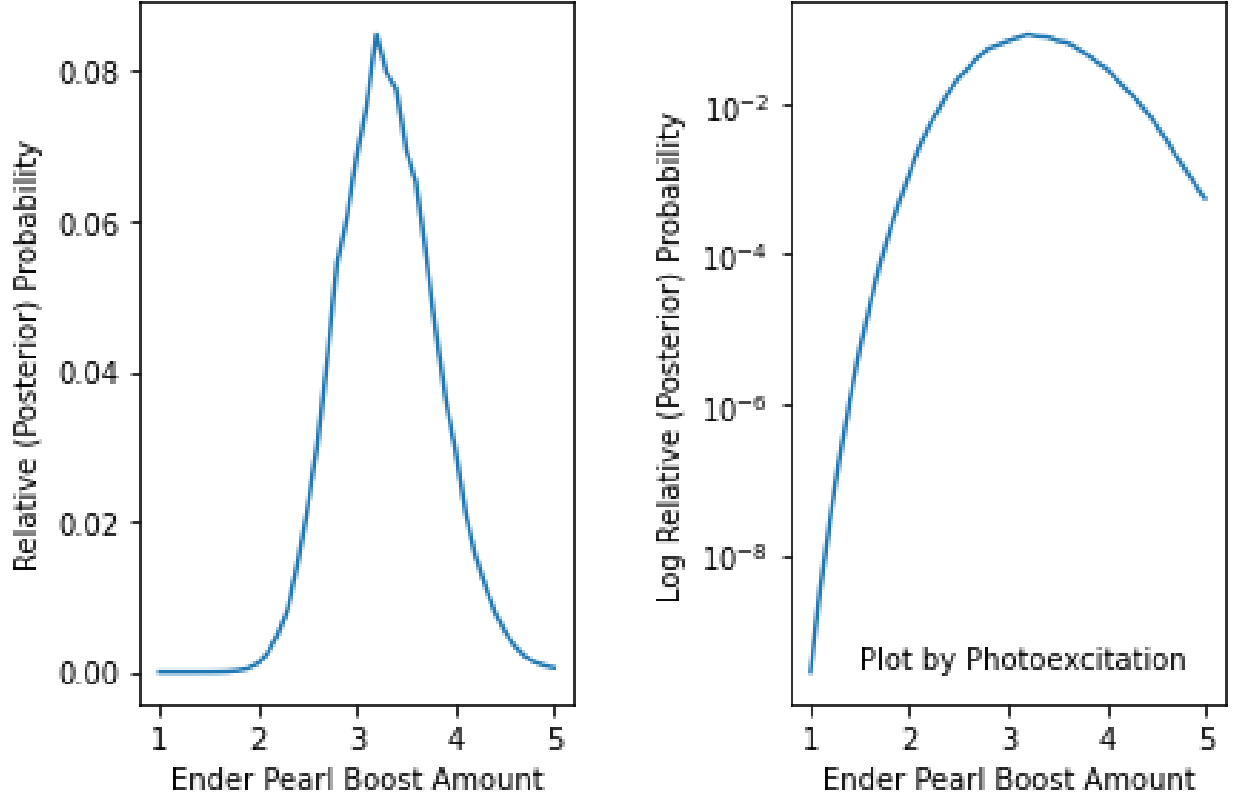


Figure 2: Bayesian probability estimate for how much the ender pearl barter probability would need to be increased in order to explain Dream's data. Note that using a probability boost in the statistical calculation does not assume that a boost was applied; the boost=1 case on the x-axis is the case where no modification was used. The fact that this is a very low probability event is not entirely surprising as Dream's data was specifically selected because it was low probability, as I discuss further in the main text. This calculation does not include removing the last attempt. This calculation suggests that the probability that the ender pearl probabilities were not boosted is about 3×10^{-10} .

6.5 Joint Probability

As blaze rods are used in conjunction with ender pearls, it makes sense to consider them together. I will implement this below after discussing another issue with the MST Report.

7 Inappropriate Correction for Sampling Bias

7.1 Inaccurate Correction for Lucky Streaks

The MST Report hypothesizes that Dream turned on modifications after the first five of eleven somewhat equivalent streams "due to a belief that, if he cheated, it was likely from the point of his return to streaming rather than from his first run." (Section 8.2) They then decide to weaken this hypothesis – in an attempt to produce a best-case scenario for Dream – and instead investigate the hypothesis that k consecutive streams of 11 were modified.

They then propose that the p-value across n streams has an upper limit of (their Equation 4)

$$p_n \leq (1 - (1 - p)^{\frac{n(n+1)}{2}}) \quad (1)$$

because there are $n(n+1)/2$ possible choices of consecutive streams. First, let us simplify this expression (and their Equation 5) by noting that all the probabilities in the paper are extremely small and it is thus an excellent approximation (far better than my factor of two precision goal) to write $(1 - (1 - p)^x) \simeq xp$. That is, choosing any substream, choosing any runner, and choosing any type of event to analyze (MST Report Sections 8.2, 8.3, 8.4 and Equations 4, 5, and 6 respectively) are all very similar corrections. Known as a Bonferroni correction, they basically say that when you want to reject the null hypothesis with probability p by trying N times, you should use a p-value of p/N . This makes sense... if you have more chances, you are more likely to experience low probability events. Note that the methodology is not in strict keeping with the premise of hypothesis testing (since typically a p-value is chosen in advance of the analysis), though that does not mean it is not meaningful.

The MST Report claims that $p_n \leq np$ places a strict Dream-benefiting upper bound on the probability because equality is only achieved if all the n tests are fully independent. As full independence is not likely, they claim $p_n < np$ and the probability is an upper bound.

However, the Bonferroni correction is not always accurate in this case because it not only assumes that all the values of p are independent, but also that they are all equal⁴. This is a very poor approximation to p-values from actual subsets because each event in the set has a probability less than one which means that subsets of different lengths will have *very* different probabilities. The lowest probability will *always* be from all 11 events.

Further, this correction does not fully account for the case where the most extreme event is chosen, as is the case here. A few examples will suffice to show that there are issues with the substream bias correction.

Lets begin with the example discussed in the MST Report as an example: getting a run of 20 heads in 100 coin tosses. At first this seems extremely unlikely as the probability of getting 20 heads in a row is $\frac{1}{2^{20}}$, just less than 1 in a million. Applying the Bonferroni correction and saying that there are 80 choices for the starting position of the 20 successful coin tosses in the string of 100 cases gives $\frac{80}{2^{20}} = 7.629 \times 10^{-5}$ or 1 in 13000. But reading over <https://mathworld.wolfram.com/Run.html> and performing a simple Monte Carlo simulation shows that it is not that simple. The actual odds come out to be about 1 in 6300, clearly better than the supposed "upper limit" calculated using the methodology in the MST Report. This is due to the facts mentioned above: 1) subsets with different p-values are harder to combine and 2) "lucky streaks" are not average randomly chosen samples, but samples that are specifically investigated because they are lucky.

Even stronger differences between numerical simulations and the proposed correction are seen when the probabilities are more extreme than the 50/50 chance of a coin toss. For example, the probability of three consecutive 1% probability events would have a p-value (from Equation 2 below) of 1.1×10^{-4} . The Bonferroni corrected probability is 8.8×10^{-4} , but a Monte Carlo simulation gives 70×10^{-4} .

⁴Technically, the Bonferroni correction should be the sum of all possible p-values, but this is difficult to calculate in practice.

These numbers serve to illustrate that **the sampling bias may not accurately accounted for** and the claim that the p-values given in the MST Report are as favorable as possible is not supported by my investigation.

On the other hand, the benefit of choosing any lucky "substreak" was actually not in keeping with their main hypothesis that Dream modified his runs at a very specific time. The other corrections (for the number of speedrunners and p-hacking) are more realistically accounted for and are discussed in detail below.

7.2 Actual Probability of Lucky Streaks

Statistics from Extreme Value Theory⁵ focuses on the probability of finding unusual events. For example, the p-value associated with getting a value of z from the product of n independent events with probability random from 0 to 1 (e.g., p-values) is:

$$\int_0^z \frac{(-\ln(x))^{n-1}}{(n-1)!} dx \quad (2)$$

When combining different independent p-values using products, I will use this Equation 2 as it is more appropriate. For the case of $n = 2$ (e.g., p-values from ender pearls and blaze rods), this equation simplifies to $z(1 - \ln z)$. That is, in addition to multiplying the two p-values together, you should also adjust the probability upwards by $(1 - \ln z)$. In this case, where the probabilities are typically very small, $-\ln z$ can be significant (around 10-50) and important to include. When testing on values, I noticed that it is possible that this result is very similar under certain conditions to Fisher's Method for combining probabilities discussed in Section 10.2.3 of the MST Report. I have not attempted to prove this, but it is good to point out that my method for combining p-values gives similar results to the method used by the MST Report.

7.3 Including all 11 streams

Dream has provided me with data on the other 5 streams. These are available at <https://drive.google.com/file/d/1Evxcv04-guI73FH5pMUJ-zEHhV-L1yuJ/view> with some of the key numbers located in the Code Snippets below. I have not confirmed the information in these data and have used them as is.

Before considering the results of these, you might find it useful to think about your personal opinion on this question. If the probabilities were modified, what was the chance that Dream did it during at this point in his speedruns? More on that later.

The ender pearl and blaze rod data for these 5 streams are uneventful: 12/356 ender pearl trades and 73/134 blaze rod drops. The cumulative binomial probabilities for these cases – even without applying the Barter Stopping Correction – are 0.86 and 0.13, e.g., these are completely consistent with chance. The analysis of just these 5 streams would show a typical outcome.

Combining all 11 streams together gives a total of 618 gold barter resulting in 54 ender pearl trades gives a naive cumulative binomial probability of 7.6×10^{-6} (without Barter Stopping Correction). For all 11 streams, there were 439 blaze kills with 284 rod drops, giving a naive cumulative binomial probability of 2×10^{-10} . Including the additional 5 "normal" streams significantly lowers the probabilities, but the combination of all 11 is still rather low probability, although the conclusion of whether this is unusual requires the additional discussion below.

For the case of ender pearls, including all the streams into my Bayesian analysis that accounts for Barter Stopping gives a probability of 3×10^{-4} for the boost=1 case when the last run is included and 2×10^{-4} when it is excluded. I take 3×10^{-4} as my best estimate. The Bayesian analysis for blaze rods would give about 10^{-6} .

Naturally, combining five "normal-looking" streams with six "extraordinary" streams leads to eleven streams that are somewhat in between. As we will see below, the probabilities associated with all eleven streams are consistent with chance, but the probabilities associated with just the last six streams are still very improbable.

In this case, it seems very natural to say "well, then the modifications must have occurred in between the fifth and sixth stream," which is one of the hypotheses put forward by the MST Report. However, as is discussed throughout this document, choosing to put a break point between the streams *after* seeing the

⁵https://en.wikipedia.org/wiki/Extreme_value_theory

probabilities would require including a correction for the bias of knowing this result. Low probability streaks are far more obvious in hindsight which leads to a temptation to associate them with incredible luck or cheating.

8 Other Corrections

As pointed out in the MST Report, since Dream was investigated because his numbers appeared lucky, an additional correction is needed to address this bias.

Given N investigations each represented by a p-value randomly drawn from 0 to 1, what is the worst p-value that you'll see? When correcting for the fact that only the worst cases (out of N possible cases) are investigated, some care must be taken. For example, there's a 1% chance that out of 100,000 random p-values to find a minimum of 10^{-7} .

Monte Carlo simulations and an investigation of extreme value statistics show that the correction for choosing the worst p-value is to multiply by the number of possible investigations. This is equivalent to the Bonferroni correction used in the MST Report.

In Section 8.3, they claim that their calculation of p is for a runner within their entire speedrunning career. This is presumably based on the argument from Section 8.2 that they have already corrected for every possible subset of streams. As I pointed out above, that correction was inaccurate. Further, that correction was based on choosing 6 of 11 livestream events from Dream, suggesting that their definition of "career" is 11 multi-hour livestream events comprising about 50 runs.

Let's instead suppose that there are 300 livestream speedruns posted per day. This is based on perusal of the recordboard at https://www.speedrun.com/mc#Any_Glitchless which shows that new records within the top 1000 runs happen about once a month, i.e., 30 per day. There are likely at least 10 times as many livestreams as there are record-holders each day, giving us 300 livestream runs per day and thus 10^5 livestream runs per year.

There are about 10^5 sets of 25 or 50 consecutive livestream runs of a specific length. That means that there's a healthy 1% chance that one of the speedrunners will experience a 10^{-7} event *chosen in advance* per year during a set of six speedruns similar to Dream.

8.1 How many random events are important?

As discussed in the MST Report, I need to use a "p-hacking correction" that acknowledges that only the most unusual random occurrences will be investigated at this level of detail. The p-hacking correction addresses the issue of focusing on only those random events that seem unusual. For example, ender pearls seemed unusual and so they were investigated as opposed to iron golems. Blaze rods were also investigated, although the reason for this choice is less clear.

The MST Report proposes that there are about 10 areas where random numbers affect the outcome at a level comparable to ender pearls and blaze rods. Dream, in coordination with other speedrunners, has identified a list of nearly 40 cases where random numbers affect the outcome comparably to blaze rods. https://docs.google.com/document/d/1izin_dl8PwuF5jFaiVwKSGBs_tfrpDj3tQdE_RwCgKM/edit?usp=sharing

As mentioned earlier, no one could possibly check every possible situation for every possible speedrunner to look for unusual cases. Suppose a speedrunner seemed to have unusual luck in, say, bartering obsidian rates. If this would precipitate an investigation similar to this one, then a speedrunner has many ways to get lucky each and this bias needs to be accounted for. This is the premise of p-hacking⁶.

If I use the 37 types of random events identified and are allowed to choose any two to combine, that leads to a p-hacking correction of $37 \times 36 \simeq 1000$ instead of the 90 used in the MST Report.⁷

Another way of handling this is to only look at ender pearls (as this was the original item that appeared unusual) and ignore blaze rod drops entirely. That would make the observed data much more plausible. Thus, to be specific, the hypothesis being tested is that two random probabilities were modified.

⁶see https://en.wikipedia.org/wiki/Data_dredging for more information

⁷Seeing the number of ways of influencing the outcome, something to consider is whether there are more clever ways to surreptitiously improve times than ender pearl bartering and blaze rod dropping.

8.2 Combined Correction

If we then ask, what is the chance that a previously unidentified "lucky" event with p-value p occurs in a leaderboard-worthy livestream per year, the answer is $p \times 10^5 \times 1000$ per year. I will focus on the last year and use a correction of 10^8 . This very large boost is a natural result of the fact that only low probability events are investigated.

It would not be hard to come up with a different correction that is also plausible. For example, you expand the list to any Minecraft speed runners in the last ten years. You could also expand the list to all the people who could have been investigated for cheating in any online competition, where the numbers obviously get much larger. Why is it so easy to change the answer? *Because the question is also changing.* When considering any Minecraft livestreamed speedrun in the last ten years, the question is "What is the probability that any runner in the Minecraft speedrunning community experienced events as rare as Dream while livestreaming in the last ten years?" When considering anyone ever accused of cheating in an online competition, the question becomes "What is the probability that anyone accused of cheating in an online competition would experience events as rare as Dream?" It is up to *you* to decide what question is important to you and then to compute your probability accordingly.

If you ask "What is the probability that anyone playing Minecraft ever had luck as good as Dream did during these 11 streams?" then the odds are very high. Another way of putting this is that Dream's luck can be described not in terms of unusual success in the game, but that out of all the Minecraft players, it was him who got lucky (in this particular way) and he got lucky while livestreaming. But remember that this cannot be counted against him specifically because he was investigated precisely because he was so lucky (like Feynman's license plate).

8.3 Comparing to Other Speedrunners

Given the probabilities and odds discussed in this document, the next step for any reader is to use this information to draw conclusions. Each reader will have a different question in mind from "Should I keep watching Dream even though he could be dishonest?" to "How much faith should I put in speedrunning leaderboards?" and many other possibilities. Many of these questions stem fundamentally from the question "Did Dream intentionally modify his probabilities?"⁸

One way of ruling out some classes of explanations is to compare Dream's results to other livestream speedrunners. For example, code glitches might affect everyone equally. Although the selection of specific streams is not discussed in detail, the comparison to other speedrunners shows that Dream's runs were highly unusual. But the fact that Dream's runs were very low probability had already been established and comparison to other runners doesn't really influence this assessment. Comparison to other runners is not necessary to establish that Dream had very low probability runs. Instead this comparison is more relevant to the *interpretation* of these low probabilities. For example, it reduces the plausibility that the low probabilities were due to some universal glitch that affects all speedrunners. As the reader is assessing the evidence, the low probability of Dream's runs and that Dream performed much better than other speedrunners **should not be considered as independent pieces of evidence** as they both are consequences of the same thing. Any lucky speedrunner chosen because they look lucky will look lucky when compared to other speedrunner streams that were chosen randomly.

9 Conclusions

If you are asking about the hypothesis that Dream was using modifications for the six streams in question, then the ender pearl barter probability was 3×10^{-10} to 3×10^{-9} depending on how you implement the stopping criterion; let's choose 10^{-10} . The blaze rod probability was 3×10^{-8} . Combining these two probabilities using Equation 2 gives 1.2×10^{-16} . Adding in the correction (by multiplying) for 100,000 possible sets of 11 streams to investigate in 1,000 different ways gives an investigation gives 10^{-8} or a 1 in 100

⁸Although I have only spent a small amount of time looking at the online discussion of all this, one hypothesis I see that may not be getting enough traction is that the modifications were present but unintentional. One version of this is that there were issues with the Random Number Generators, but the MST Report concludes that this is extremely unlikely. I have enough experience with code to say that completely unexpected consequences can happen, even after poring over the code in detail.

million chance. That is, there is a 1 in 100 million chance that a livestream in the Minecraft speedrunning community got as lucky this year on two separate random modes as Dream did in these six streams. That is extraordinarily low, though not nearly as low (by a factor of 75000) as concluded by the MST Report (1 in 7.5 trillion). The main things that increased the probability are: 1) using a Barter Stopping criterion (factor of about 100) and 2) using 100 times as many livestreams and 10 times as high a p-hacking correction, for which I have provided specific justification.

If you are asking about the hypothesis that Dream was using modifications for all eleven streams, the probabilities are much higher because the other five streams had more typical results. The ender pearl probability goes up to 3×10^{-4} and the blaze rod probability goes up to 10^{-6} . Combining these gives 7×10^{-9} and adding the 10^8 boost gives 0.7 or 1 in 2. Note that my corrections are designed for low p-values, so this may not be fully accurate, but this inaccuracy would not affect the conclusion that this case is completely consistent with expectations. That is, an investigation of all the similar Minecraft livestreams that picked a runner who had unusual luck in two different ways would produce results as unusual as Dream's in these 11 streams. Note that for speedrunners to reach high positions on the leaderboard requires excellent skill *and* luck.

These answers are extremely different, which is unsurprising because the ender pearl and blaze rod success rate is very different between the first five and last six streams. How should you decide between the case with eleven streams and the case with six streams? It depends on what you think the probability is that Dream would make a modification at that point (as compared to any other point) without being influenced by the actual probabilities. It was a natural breaking point in the timeline of streams independent of the fact that it was probabilistically extremely different, which argues for the six-stream hypothesis. If you allow the streak of streams/runs to be any length up to N (instead of choosing 6 or 11 in advance), then another correction of N^9 should be included. Using $N \simeq 10$ gives a corrected probability of 1 in 10 million. This does not account for the fact that "lucky streaks" should be treated somewhat differently which would increase the odds, potentially up to 1 in a million.

So if you think "if Dream would have chosen to modify his numbers then this is the only place within the eleven stream set that Dream would have modified them", then you should lean toward the 1 in 100 million case. If you think Dream could have chosen to modify his numbers in between any stream, then these odds should come down substantially to 1 in a 10 million. If you think that if Dream modifying things, he would only have done it at the beginning of all eleven streams in question, then the data show no statistically significant evidence that Dream was modifying the probabilities, given that he was investigated after it was noticed that he was lucky.

Since the eleven-stream probability is so much higher, even if you think that (independent of the probabilities calculated after seeing the streams) there is a 100-to-1 chance Dream modified before the final six streams instead of before all eleven streams, the six stream case provides a negligible correction and the probability becomes just 1/100. That is, external evidence that the probabilities were modified at this specific point would be needed to produce a significant probability of cheating.

Even in the worst case, the probabilities are not so extreme as to completely rule out any chance that Dream used the unmodified probabilities. If you have independent high-probability reasoning to suppose that the game was modified by Dream before his final six runs, then the low probability of that hypothesis even after correcting for other biases suggests an alternative explanation. There are reasonable explanations for Dream's ender pearl and blaze rod probability, potentially including extreme "luck", but the validity and probability of those explanations depend on explanations beyond the scope of this document. One alternative explanation is that Dream (intentionally or unintentionally) cheated, though I disagree that the situation suggests that this is an unavoidable conclusion. In any case, the conclusion of the MST Report that there is, at best, a 1 in 7.5 trillion chance that Dream did not cheat is too extreme for multiple reasons that have been discussed in this document.

⁹The MST Report computes this in a different way: choosing the number of 11-stream livestreamers and then choosing any of $11 \cdot (11+1)/2$ subsets from these streams. In addition to the possible issues mentioned above, this correction is pretty strongly dependent on the somewhat arbitrary choice of 11 (which is potentially relevant to Dream, but may not be universal). I instead propose that you take consider all sets of consecutive livestreams of a certain length, which leads to a correction of the number of livestreams times the number of plausible lengths

A Code Snippets

Some code snippets are shown here for reference. I used python in the form of an ipython notebook.

```
import numpy
import matplotlib.pyplot as plt
from scipy.stats import binom
from matplotlib.lines import Line2D
import copy

# number of gold barbers needed to get numneeded ender pearls
# assuming that after numneeded (10) ender pearls are obtained, trading stops
# uses an algorithm that includes a random number of ender pearls obtained per trade
# (though this does not matter for 10 ender pearls
# since only 2 trades are needed for this goal

# to test the hypothesis that the probability of ender pearls was somehow boosted
# calculate the number of gold barbers needed for a variety of ender pearl
# probabilities from the nominal 20/423 (probpearlboost=1) up to 100/423
# (probpearlboost=5), representing a uniform prior of 1-5 for this boost probability

# note, this simulation can take several minutes

prob_pearls=20.0/423.0
num_prob_pearl_boost=41
prob_pearl_boost_arr=numpy.linspace(1,5,num_prob_pearl_boost)
num_monte_carlo=1000000 # number of monte carlo simulations
num_pearls_needed=10

# stores the number of gold needed in each simulation for each boost
gold_needed_arr=numpy.zeros([num_monte_carlo,num_prob_pearl_boost])

# stores the number of successful pearl trades needed
trades_needed_arr=numpy.zeros([num_monte_carlo,num_prob_pearl_boost])

# loop over each boost and each simulation
for iboost in range(num_prob_pearl_boost):
    for imc in range(num_monte_carlo):
        # reset the simulation
        current_pearls=0
        current_gold=0
        current_trades=0
        # trade until the number of pearls obtained
        while current_pearls < num_pearls_needed:
            # do one gold barter
            current_gold=current_gold+1
            # check if this barter leads to an ender pearl trade
            # using boosted probability
            if numpy.random.uniform()<prob_pearls*prob_pearl_boost_arr[iboost]:
                current_trades=current_trades+1
                # give between 4-8 pearls
                # approximating the observed distribution
                current_pearls=current_pearls+numpy.round(
                    4*numpy.random.uniform()+0.5) + 3
```

```

gold_needed_arr[imc, iboost]=current_gold
trades_needed_arr[imc, iboost]=current_trades

# now take the simulation results and turn them into a probability of
# getting 10 ender pearls given a specific number of gold bartered and
# a specific probability boost

max_gold=500 # maximum number of gold barters in the array
prob_this_gold_arr=numpy.zeros([max_gold, num_prob_pearl_boost])
for igrid in range(num_prob_pearl_boost):
    for this_gold in range(max_gold):
        prob_this_gold_arr[this_gold, igrid]=

        numpy.sum(gold_needed_arr[:, igrid]==this_gold)/num_monte_carlo

# data from Dream trades on 11 streams of interest
# see MST Report Appendix A and http://bombch.us/DPPU
dream_trades= [22,5,24,18,4,1,7,12,26,8,5,20,2,13,10,10,21,20,10,3,
18,3,27,4,13,5,35,70,11,7,24,34,7,15,10,1,40,50,5]
dream_successes=[3,2,2,2,0,1,2,5,3,2,2,2,0,1,2,2,2,2,2,1,
2,2,2,0,0,1,1,2,0,1,0,0,0,0,0,0,0,3,2,0]
dream_goalof12= [1,0,0,1,0,0,0,0,1,1,0,0,0,0,1,1,1,1,1,0,
1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0]
dream_goalof10= [1,1,1,1,0,0,1,1,1,1,1,1,0,0,1,1,1,1,1,0,
1,1,1,0,0,0,0,1,0,0,0,0,0,0,0,0,1,1,0]

# data from Dream trades on 6 streams of interest
# see MST Report Appendix A and http://bombch.us/DPPU
#dream_trades= [22,5,24,18,4,1,7,12,26,8,5,20,2,13,10,10,21,20,10,3,18,3]
#dream_successes=[3,2,2,2,0,1,2,5,3,2,2,2,0,1,2,2,2,2,2,1,2,2]
#dream_goalof12= [1,0,0,1,0,0,0,0,1,1,0,0,0,0,1,1,1,1,1,0,1,0]
#dream_goalof10= [1,1,1,1,0,0,1,1,1,1,1,1,0,0,1,1,1,1,1,0,1,1]

# probability calculation for each individual trade
# "goalof10" trades use Barter Stopping Probability
# other trades use Binominal Probability

this_trade_prob=numpy.zeros([len(dream_trades), num_prob_pearl_boost])

for iboost in range(num_prob_pearl_boost):
    for this_trade in range(len(dream_trades)):
        if dream_goalof12[this_trade] == 1:
            this_trade_prob[this_trade, iboost]=

            prob_this_gold_arr[dream_trades[this_trade], iboost]
        else:

            # probability when trades = successes is prob^successes
            if dream_trades[this_trade]==dream_successes[this_trade]:

```

```

        this_trade_prob[this_trade, iboost]=
        (prob_pearls*prob_pearl_boost_arr[iboost])** (dream_successes[this_trade])
    else:
        this_trade_prob[this_trade, iboost]=
        binom.pmf(dream_successes[this_trade],
        dream_trades[this_trade], prob_pearls*prob_pearl_boost_arr[iboost])

# allow for ignoring the last barter to correct for optional stopping
ignore_last_barter = True
if ignore_last_barter:
    last_barter_correction=-1
else:
    last_barter_correction=0

total_prob=numpy.product(this_trade_prob[0:len(this_trade_prob)
+last_barter_correction, :], axis=0)
print(total_prob/numpy.sum(total_prob))

```